10/079,999 #3

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application: ２００１年１１月２８日

出願番号
Application Number: 特願２００１－３６２８５８

ST.10/C]: [ＪＰ２００１－３６２８５８]

出願人
Applicant(s): 株式会社日立製作所

CERTIFIED COPY OF
PRIORITY DOCUMENT

２００２年 ２月 ５日

特許庁長官
Commissioner,
Japan Patent Office 及川耕造

【書類名】　特許願

【整理番号】　KN1350

【提出日】　平成13年11月28日

【あて先】　特許庁長官殿

【国際特許分類】　G06F 13/00

【発明者】

【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内

【氏名】　水野　陽一

【発明者】

【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内

【氏名】　松並　直人

【発明者】

【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内

【氏名】　味松　康行

【発明者】

【住所又は居所】　神奈川県小田原市中里３２２番地２号　株式会社　日立製作所　ＲＡＩＤシステム事業部内

【氏名】　▲高▼本　賢一

【特許出願人】

【識別番号】　000005108

【氏名又は名称】　株式会社　日立製作所

【代理人】

【識別番号】　100093492

【弁理士】

【氏名又は名称】　鈴木　市郎

【電話番号】　03-3591-8550

【選任した代理人】

【識別番号】　100078134

【弁理士】

【氏名又は名称】　武　顕次郎

【手数料の表示】

【予納台帳番号】　113584

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【プルーフの要否】　要

【書類名】　　明細書

【発明の名称】　ディスクアレイシステム及びコントローラ間での論理ユニットの引き継ぎ方法

【特許請求の範囲】

【請求項１】　１または複数の計算機と、それぞれが専用のキャッシュを備えた複数のコントローラ及び複数のディスク装置を備えた前記計算機により使用されるディスクアレイ装置とにより構成されるディスクアレイシステムにおいて、前記コントローラは、ディスク装置に作成される論理ユニットの構成情報を管理する構成管理手段を備え、該構成管理手段は、前記論理ユニットの構成情報を書き替えることにより前記コントローラが担当する任意の論理ユニットを別の任意のコントローラに引き継ぐ処理を行うことを特徴とするディスクアレイシステム。

【請求項２】　前記コントローラ間での論理ユニットの引き継ぎの際、対象論理ユニットと移行先コントローラとが指定されることを特徴とする請求項１記載のディスクアレイシステム。

【請求項３】　前記コントローラ間での論理ユニットの引き継ぎの際、移行先コントローラでの接続ポート番号または前記計算機から認識される論理ユニット番号が変更されることを特徴とする請求項１記載のディスクアレイシステム。

【請求項４】　前記コントローラ間での論理ユニットの引き継ぎの際、前記キャッシュ上の対象論理ユニットのデータを前記ディスク装置に書き込むことを特徴とする請求項１記載のディスクアレイシステム。

【請求項５】　前記キャッシュ上のデータの書き込みは、書き込みを指示するコマンドによって行われることを特徴とする請求項４記載のディスクアレイシステム。

【請求項６】　前記構成管理手段は、移行先コントローラを記憶しておくことにより、自動的な論理ユニットの引き継ぎ行うことを特徴とする請求項１記載のディスクアレイシステム。

【請求項７】　１または複数の計算機と、それぞれが専用のキャッシュを備えた複数のコントローラ及び複数のディスク装置を備えた前記計算機により使用

されるディスクアレイ装置とにより構成されるディスクアレイシステムのコントローラ間での論理ユニットの引き継ぎ方法において、前記コントローラは、ディスク装置に作成される論理ユニットの構成情報を管理し、前記論理ユニットの構成情報を書き替えることにより前記コントローラが担当する任意の論理ユニットを別の任意のコントローラに引き継ぐ処理を行うことを特徴とするコントローラ間での論理ユニットの引き継ぎ方法。

【請求項８】　前記コントローラ間での論理ユニットの引き継ぎの際、前記キャッシュ上の対象論理ユニットのデータを前記ディスク装置に書き込むことを特徴とする請求項７記載のコントローラ間での論理ユニットの引き継ぎ方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、ディスクアレイシステム及びコントローラ間での論理ユニットの引き継ぎ方法に係り、特に、複数のコントローラのそれぞれが専用のキャッシュを備えたディスクアレイシステム及びコントローラ間での論理ユニットの引き継ぎ方法に関する。

【０００２】

【従来の技術】

ディスクアレイシステムに関する従来技術として、複数のディスク装置のそれぞれに対応して設けられたコントローラと、各コントローラからアクセス可能な共有キャッシュとを備えたディスクアレイシステムが知られている。

【０００３】

図１０は共有キャッシュを備えた従来技術によるディスクアレイシステムの構成を示すブロック図であり、図１０を参照して従来技術について説明する。図１０において、１００はディスクアレイ装置、２００ｘ（ｘ＝ａ……ｎ）はコントローラ、３００は共有キャッシュ、４００は共有メモリ、５００ｘ（ｘ＝ａ……ｎ）はディスク装置、６００は共通バスである。

【０００４】

図１０に示すディスクアレイ装置１００は、図示しない他の計算機等と共に、

ディスクアレイシステムを構成するものであり、ディスク装置を説明する複数のコントローラ２００ｘと、入出力データを格納する共有キャッシュ３００と、構成情報等を格納する共有メモリ４００と、複数のディスク装置５００ｘと、これらを接続する共通バス６００とにより構成されている。共有キャッシュ３００及び共有メモリ４００は、どのコントローラ２００ｘからも共通バス６００を通してアクセス可能である。共有キャッシュ３００は、全てのコントローラ２００ｘからアクセスが集中するため、高価な大容量のものが用意される。そして、各コントローラ２００ｘは、それぞれ担当するディスク装置５００ｘが予め決められている。

【０００５】

前述した構成の従来技術は、あるコントローラ２００ｘが担当するディスク装置５００ｘを別のコントローラ２００ｘに切り替える場合、ディスク装置５００ｘに関する入出力データが共有キャッシュ３００上に置かれているため、移行先のコントローラが共有キャッシュ上のデータをそのまま使用して、ディスク装置５００ｘとコントローラ２００ｘとの関係を変更することができる。また、引継ぎに必要なディスク装置の構成情報等は共有メモリ４００に格納されており、移行先コントローラは、共有メモリ４００を参照して、対象ディスク装置の制御を開始すればよい。また、移行元コントローラは、これまで制御していたディスク装置５００ｘに対するアクセス要求を拒否し、ディスク装置５００ｘの管理を止めるだけでよい。これにより、前述の共有キャッシュを備えた従来技術によるディスクアレイシステムは、あるコントローラが担当するディスク装置を切り替えることにより、コントローラが担当するディスク装置内の論理ユニットに対するアクセス量を均等化してコントローラの負荷分散を行うことができる。

【０００６】

また、前述の従来技術は、ディスクを移設し、物理位置を移動することによって、負荷分散を図ることが可能である。この場合、コントローラは、対象となるディスク装置へのアクセスを停止し、ディスク装置をシステムから完全に切り離し、次に、切り離したディスク装置が物理的に異なる位置に差し込まれたとき、ディスク装置を再認識して処理を開始する。また、前述の従来技術は、ディスク

を異なるディスクアレイ装置に移動することも可能である。これにより、前述の従来技術は、再認識されたディスク装置を新しい別のコントローラに割り当て、制御するコントローラを切替えることにより、前述と同様に、負荷分散を行うことができる。

【０００７】

【発明が解決しようとする課題】

前述した従来技術は、共有キャッシュを備えたディスクアレイ装置に適用した場合に好適なものであったが、各コントローラが独立した専用キャッシュを備えたディスクアレイ装置に適用することができなかった。

【０００８】

各コントローラが独立した専用キャッシュを備えている分散キャッシュ環境のディスクアレイシステムは、各コントローラがそれぞれ専用キャッシュを使用して個々のボリュームを担当することにより、共有キャッシュ環境の場合に生じていたようなアクセスの集中を回避し、コストパフォーマンス及びスケーラビリティを上げることができる。

【０００９】

しかし、各コントローラが独立した専用キャッシュを備えている分散キャッシュ環境のディスクアレイシステムは、あるコントローラが担当するディスク装置を切り替えようとする場合、移行先コントローラが対象となるボリュームの入出力データを保持していないため、コントローラをそのまま切り替えることができないという問題点を有している。

【００１０】

このような問題は、全てのコントローラのデータキャッシュに常に同じ内容を書き込んでおけば解決することができ、共有キャッシュの場合と同様にボリュームを引き継ぐようにすることが可能であるが、各コントローラのキャッシュに担当でないボリュームのデータも書き込むことになり、キャッシュ容量を圧迫し、分散キャッシュのメリットが失われることになるという問題点を生じる。

【００１１】

また、各コントローラが独立した専用キャッシュを備えているディスクアレイ

システムは、ディスクの移設の自動的な切り替え及びダイナミックな切り替えを行うことができないという問題点を有している。すなわち、分散キャッシュ環境のディスクアレイシステムは、ディスクをシステムから切り離す作業と、切り離したディスクの物理位置を変更する作業と、ディスクを再認識させる作業とが必要であり、人手を介してこれらの作業を行わなければならず、この間、対象ディスクへのアクセスを完全に停止しなければならず、しかも、この場合、ディスクを元のコントローラに戻すことは考慮されていなかった。

【００１２】

前述したように、各コントローラが独立した専用キャッシュを備えている従来技術のディスクアレイシステムは、システムを停止せずにダイナミックに担当コントローラを切り替えることができないという問題点を有している。

【００１３】

本発明の目的は、前述した従来技術の問題点を解決し、システムを中断させることなく任意のコントローラ間で任意のボリュームをダイナミックに引継ぐことを可能とした、各コントローラが独立した専用キャッシュを備え、それぞれ個々のボリュームを担当しているディスクアレイシステム及びコントローラ間での論理ユニットの引き継ぎ方法を提供することにある。

【００１４】

【課題を解決するための手段】

本発明によれば前記目的は、１または複数の計算機と、それぞれが専用のキャッシュを備えた複数のコントローラ及び複数のディスク装置を備えた前記計算機により使用されるディスクアレイ装置とにより構成されるディスクアレイシステムにおいて、前記コントローラが、ディスク装置に作成される論理ユニットの構成情報を管理する構成管理手段を備え、該構成管理手段は、前記論理ユニットの構成情報を書き替えることにより前記コントローラが担当する任意の論理ユニットを別の任意のコントローラに引き継ぐ処理を行うことにより達成される。

【００１５】

また、前記目的は、１または複数の計算機と、それぞれが専用のキャッシュを備えた複数のコントローラ及び複数のディスク装置を備えた前記計算機により使

用されるディスクアレイ装置とにより構成されるディスクアレイシステムのコントローラ間での論理ユニットの引き継ぎ方法において、前記コントローラが、ディスク装置に作成される論理ユニットの構成情報を管理し、前記論理ユニットの構成情報を書き替えることにより前記コントローラが担当する任意の論理ユニットを別の任意のコントローラに引き継ぐ処理を行うことにより達成される。

【００１６】

【発明の実施の形態】

以下、本発明によるディスクアレイ装置の実施形態を図面により詳細に説明する。

【００１７】

図１は本発明の第１の実施形態によるディスクアレイシステムの構成を示すブロック図、図２は構成管理手段が備える構成情報テーブルの構成例を示す図である。図１、図２において、１はディスクアレイ装置、２ｘ（ｘ＝ａ、ｂ、……、ｎ）は計算機、３は経路制御装置、４は管理コンソール、５ｘ（ｘ＝ａ、ｂ、……ｎ）はチャネルパス、６ｘ（ｘ＝ａ、ｂ、……、ｎ）はＬＡＮ（Local Area Network）、７は通信手段、１０はデバイスネットワーク、１１ｘ（ｘ＝ａ、ｂ、……、ｎ）はコントローラ、１２ｘ（ｘ＝ａ、ｂ、……、ｎ）はディスク装置、３１はパス変更手段、４１は管理ユーティリティ、１１１ｘ（ｘ＝ａ、ｂ、……、ｎ）はデータキャッシュ、１１２ｘ（ｘ＝ａ、ｂ、……、ｎ）は構成管理手段、１２１ａ～１２１ｆはＬＵ（Logical Unit）である。

【００１８】

本発明の第１の実施形態は、図１に示すように、ディスクアレイ装置１と、ディスクアレイ装置１を使用する複数の計算機２ｘと、全ての計算機２ｘとディスクアレイ装置１と管理コンソール４を相互に接続する経路制御装置３と、ディスクアレイ装置１を管理するための管理コンソール４とが、チャネルパス５ｘにより相互に接続されて構成されている。図示例では、チャネルパス５ｘして、ファイバチャネルを使用することとする。また、図示実施形態は、計算機２ｘと管理コンソール４とが通信を行うためのＬＡＮ６ｘと、ディスクアレイ装置１と管理コンソール４とが通信するための通信手段７とが備えられる。

【００１９】

経路制御装置３には、計算機２ｘからディスクアレイ装置１へのアクセスパスを変更するためのプログラムであるパス変更手段が備えられる。図１に示す例では、パス変更手段３１を経路制御装置３上に配置するものとしているが、同様の機能を計算機２ｘ上に配置するようにすることもできる。

【００２０】

管理コンソール４には、ディスクアレイ装置１内のディスク装置１２ｘ内部のＬＵ構成の表示や、システムの管理者がディスクアレイ装置１のＬＵ設定等を行うために使用される管理ユーティリティ４１が備えられる。図１に示す例では、管理ユーティリティ４１を管理コンソール４上に配置するものとしているが、管理ユーティリティ４１を計算機２ｘや、ディスクアレイ装置１に配置するようにすることもできる。

【００２１】

ディスクアレイ装置１は、複数のコントローラ１１ｘと、複数のディスク装置１２ｘとが、デバイスネットワーク１０で接続されて構成される。この構成により、任意のコントローラから任意のディスク装置へのアクセスが可能である。デバイスネットワーク１０は、どのようなインタフェースによって構成してもよいが、接続性に優れたファイバチャネル等の利用が好適であり、また、総合的なスループットを高めるため、１つあるいは複数のスイッチで構成することも可能である。

【００２２】

コントローラ１１ｘは、計算機２から書き込まれたデータや、ディスク装置１２ｘから読み出されたデータを一時的に格納する入出力データ格納用のデータキャッシュ１１１ｘと、ディスクやＬＵの構成を管理する構成管理手段１１２ｘとを備えて構成される。ここで、ＬＵ（Logical Unit）とは、ディスクアレイ装置１内部に設けた仮想的な論理ボリュームのことであり、計算機２とディスクアレイ装置１とを接続するインターフェイスの１つのプロトコルであるＳＣＳＩ（Small Computer System Interface）の仕様において定義された名称である。また、ＬＵを識別するための番号のことを論理ボリューム番号（ＬＵＮ：Logical Unit

Number)と呼ぶ。

【００２３】

ディスクアレイ装置１の内部で定義したＬＵを特に内部論理ボリューム（内部ＬＵ）と呼び、ディスクアレイ装置１は、内部ＬＵを管理するため、ＬＵに０から始まる整数でシリアル番号を付ける。この番号を内部論理ボリューム番号（内部ＬＵＮ）と呼ぶ。計算機２ｘは、ＬＵを検出する際にＬＵＮを０から順にサーチし、ある番号が存在しない場合それ以降のサーチを行わない場合がある。このため、内部ＬＵＮをそのまま計算機２ｘに割り当てるのではなく、計算機２ｘが認識できるＬＵＮに変換して割り当てる必要がある。このようにして各計算機２ｘから認識されるＬＵＮを外部論理ボリューム番号（外部ＬＵＮ）と呼び、内部ＬＵＮとは区別される。

【００２４】

論理ボリュームＬＵ１２１ｘ（１２１ａ、１２１ｂ、……、１２１ｎ）は、それぞれディスク装置１２ｘ上に作成される。なお、ＬＵ１２１ｘは、どのようなＲＡＩＤ構成であってもよい。論理ボリュームＬＵの作成時に、該当するＬＵの制御を担当するコントローラが割り当てられる。１つのコントローラに対して複数のＬＵを割り当てることも可能である。

【００２５】

図１に示す構成を有するディスクアレイシステムは、各コントローラ１１ｘが専用のデータキャッシュ１１１ｘを備え、個々のＬＵの制御を担当することにより、特定のデータキャッシュへのアクセスの集中を防止することができ、共有キャッシュを備えるシステムに比べてコストパフォーマンス及びスケーラビリティを高めることができる。

【００２６】

構成管理手段１１２ｘが備える構成情報テーブル１１２１ｘの一例を図２に示している。構成情報テーブル１１２１ｘは、各ＬＵ１２１ｘの構成情報を管理するためのテーブルであり、内部ＬＵＮ、外部ＬＵＮ、ポート番号、コントローラ番号、ブロック数、ＲＡＩＤグループ番号、ＲＡＩＤレベル、物理アドレス情報等が格納される。なお、この構成情報テーブル１１２１ｘは、ディスク装置１２

ｘの予め定められた位置に格納されていてよい。

【００２７】

前述において、ポート番号は、コントローラ１１ｘが備えるファイバチャネル接続ポートのうち、そのＬＵが使用できるポートの識別番号を表す。コントローラ番号は、ディスクアレイ装置１内のコントローラの識別番号であり、その中のデフォルトコントローラ番号は、本来どのコントローラがそのＬＵを担当すべきかを表し、カレントコントローラ番号は、現在制御を担当しているコントローラを表す。ブロック数は、各ＬＵの論理ブロック数を表し、これにより、各ＬＵのサイズを知ることができる。

【００２８】

ディスクアレイ装置１は、その内部で、同一ＲＡＩＤレベルのＬＵをグループ化して管理している場合がある。ＲＡＩＤグループ番号、ＲＡＩＤレベルは、その場合の各ＬＵが属するＲＡＩＤグループの識別番号及びそのＲＡＩＤレベルを表す。物理アドレス情報は、各ＬＵの論理アドレスに対する物理的なディスク位置情報である。

【００２９】

図３はコントローラ相互間でＬＵの引き継ぎを行う場合の処理動作を説明するフローチャートであり、次に、コントローラ間のＬＵの引き継ぎについて説明する。

【００３０】

（１）管理コンソール４を操作する作業者は、移行対象となる全ＬＵのＬＵＮと、移行先コントローラとを指定する。その際、移行先コントローラで使用するポート番号、外部ＬＵＮを変更したい場合は、それぞれポート番号、外部ＬＵＮも併せて指定する（ステップ８０１）。

【００３１】

（２）管理ユーティリティ４１は、指定された情報に基づいて通信手段７を介してディスクアレイ装置１にＬＵの引き継ぎ指示を発行する（ステップ８０２）。

【００３２】

（３）移行元コントローラ１１ｘの構成管理手段１１２ｘは、ＬＵの引き継ぎ指

示を受け取ると、対象ＬＵの構成情報を移行先コントローラ１１ｘへ移行する。具体的には、移行元コントローラ１１ｘの構成管理手段１１２ｘが構成情報テーブル１１２１ｘの対象ＬＵの欄をディスクの予め決められた位置へ書き込み、移行先コントローラ１１ｘの構成管理手段１１２ｘがそれを読み込むことにより、対象ＬＵの構成情報を移行する。ポート番号、外部ＬＵＮを変更する場合、移行元コントローラ１１ｘの構成管理手段１１２ｘが、構成情報テーブル１１２１ｘを変更し、変更後の内容をディスクに書き込む。構成情報の移行は、デバイスネットワーク１０を経由して送信することにより行われる。また、図示しない専用線を経由で直接送信するようにすることも可能である。また、全コントローラ１１ｘの構成管理手段１１２ｘが、予め全ＬＵの構成情報を構成情報テーブル１１２１ｘに格納しておくこともちろん可能である。その場合、構成情報の移行を省略することができ、ポート番号、外部ＬＵＮ、コントローラ番号のみを変更すればよい（ステップ８０３）。

【００３３】

（４）次に、移行元コントローラ１１ｘは、データキャッシュ１１１ｘ上にある対象ＬＵのデータを全てデステージする。デステージとは、データキャッシュ上のデータをディスクに書きこむ処理のことである。この間、計算機２ｘからのライトアクセスは全てライトスルーとする。これにより、対象ＬＵのデータをデータキャッシュ上から全て掃き出し、ディスク上のＬＵの内容を整合性が保たれた状態にすることができる。なお、デステージの処理は、管理コンソール４からの書き込みコマンドに基づいて行うようにすることもできる（ステップ８０４）。

【００３４】

（５）移行元コントローラ１１ｘは、対象ＬＵのデステージが完了すると、管理ユーティリティ４１にデステージ完了を通知し、デステージ完了の通知を受けた管理ユーティリティ４１は、パス変更手段３１へパス切り替えを指示する（ステップ８０５）。

【００３５】

（６）パス切り替えの指示を受けたパス変更手段３１は、対象ＬＵのアクセスを一時保留する。これはＬＵ引継ぎ中の過渡状態でのアクセスを抑制するためであ

る（ステップ８０６）。

【００３６】

（７）次に、パス変更手段３１は、対象ＬＵのアクセスパスを移行元コントローラから移行先コントローラへ切り替える。これにより、対象ＬＵへのフレームは、全て移行先コントローラへ送信されることになる（ステップ８０７）。

【００３７】

（８）次に、対象ＬＵの制御を行うコントローラを移行先コントローラに切り替える。移行元コントローラは、対象ＬＵに対するアクセスを完全に停止し、移行先コントローラは対象ＬＵの制御を開始する（ステップ８０８）。

【００３８】

（９）コントローラの切り替えが完了すると、パス変更手段３１は、対象ＬＵのアクセスを再開させる（ステップ８０９）。

【００３９】

前述した処理により、ＬＵの引き継ぎが完了する。前述のような処理により、キャッシュ上の対象ＬＵのデータをデステージして、ディスクの整合性を保つことができ、各コントローラが独立した専用キャッシュを備えた場合においても、システムを中断させることなく任意のコントローラ間でボリュームを引継ぐことができる。

【００４０】

図４は図３により説明したＬＵ引き継ぎの具体的な例を説明する図、図５は引き継ぎ前後の構成テーブルの例を示す図であり、次に、図４、図５を参照して、ＬＵ引き継ぎの具体例と構成情報テーブルの書き換えとについて説明する。

【００４１】

いま、ある計算機２ｘが、図４（ａ）に示すように、コントローラ０$^{\#}$ １１ａに割り当てられているＬＵ０$^{\#}$ １２１ａ～ＬＵ３$^{\#}$ １２１ｄを使用しているものとする。そして、コントローラ１$^{\#}$ １１ｂを追加し、図４（ｂ）に示すように、ＬＵ２$^{\#}$ １２１ｃ、ＬＵ３$^{\#}$ １２１ｄの制御をコントローラ１$^{\#}$ １１ｂに切り替えることとする。このとき、外部ＬＵＮは変更せず、ポート番号のみを変更するものとする。また、簡単のため、全構成管理手段１１２ｘが全ＬＵの情報を構成

情報テーブル１１２１ｘに格納しているものとする。

【００４２】

前述の場合、引き継ぎ前の構成情報テーブル１１２１ｘは、図５（ａ）に示すような状態となっている。なお、ブロック数、ＲＡＩＤグループ番号、ＲＡＩＤレベル、物理アドレス情報等は引き継ぎによって変更されることはないので、ここでは図示を省略している。

【００４３】

図５（ａ）から全てのＬＵに対して、デフォルトコントローラ番号及びカレントコントローラ番号に、現在の担当コントローラである“０”が格納されており、また、ＬＵ０及びＬＵ１は、ポート番号０を、ＬＵ２及びＬＵ３はポート番号１を使用していることが判る。

【００４４】

管理コンソール４を操作する作業者は、移行対象であるＬＵ２$^{\#}$ １２１ｃ、ＬＵ３$^{\#}$ １２１ｄを選択して、移行先のコントローラ１$^{\#}$ １１ｂ及び移行先コントローラ１$^{\#}$ １１ｂで使用するポート番号“０”を指定する。指定された情報に基づいて、構成情報テーブル１１２１ｘが変更される。この結果、変更後の構成情報テーブル１１２１ｘは、図５（ｂ）に示すような状態になる。図５（ｂ）からＬＵ２$^{\#}$ 及びＬＵ３$^{\#}$ のポート番号が“０”に、デフォルトコントローラ番号及びカレントコントローラ番号が“１”に変更されたことが判る。これにより、ＬＵ２$^{\#}$ １２１ｃ、ＬＵ３$^{\#}$ １２１ｄの制御をコントローラ１$^{\#}$ １１ｂに切り替えることができる。

【００４５】

前述した本発明の第１の実施形態によれば、データキャッシュ上の対象ボリュームのデータをディスクに掃き出し、ディスクの整合性を保つことにより、各コントローラが独立した専用キャッシュを備えた場合においても、システムを中断させることなく任意のコントローラ間で任意のボリュームを引き継ぐことができ、システムにおける各コントローラの負荷分散を図ることができるという効果を得ることができる。

【００４６】

図６は本発明の第２の実施形態によるディスクアレイシステムの構成を示すブロック図である。図６において、２０はテープ装置であり、他の符号は図１の場合と同一である。

【００４７】

図６に示す本発明の第２の実施形態における図１に示した第１の実施形態との相違点は、計算機２ｎにテープ装置２０が接続されていることと、各計算機２ｘが各コントローラ１１ｘに直接接続されていることとである。もちろん、第１の実施形態の場合と同様に、計算機２ｘとコントローラ１１ｘとを経路制御装置を介して接続してもよいが、第２の実施形態は、パス変更手段を不要とすることができる。また、計算機２ｘとコントローラ１１ｘとは、１対１に対応している必要はなく、１つのコントローラに対して、複数の計算機を接続することも可能である。

【００４８】

図６に示す本発明の第２の実施形態は、データ二重化機能を備え、計算機２ｎを、ディスクアレイ装置１内のボリュームのバックアップを取得するために使用す一般にバックアップサーバと呼ぶ計算機としたものである。本発明の第２の実施形態は、バックアップ処理に限らず、バッチ処理やデータマイニング等の様々な処理に適用可能であるが、ここでは、本発明をバックアップ処理に適用する場合を想定して説明する。

【００４９】

図６に示すディスクアレイシステムにおいて、コントローラ１１ａには、ＬＵ０$^{\#}$ １２１ｇ及びＬＵ１$^{\#}$ １２１ｈが割り当てられ、コントローラ１１ｂには、ＬＵ２$^{\#}$ １２１ｉ及びＬＵ３$^{\#}$ １２１ｊが割り当てられているものとする。そして、ＬＵ１$^{\#}$ １２１ｈ、ＬＵ３$^{\#}$ １２１ｊは、それぞれ、ＬＵ０$^{\#}$ １２１ｇ、ＬＵ２$^{\#}$ １２１ｉの複製であり、ＬＵ０$^{\#}$ １２１ｇ、ＬＵ２$^{\#}$ １２１ｉと同一の内容のデータが格納されている。このようにディスクアレイ装置内にボリュームの複製を作成する機能を一般にデータ二重化機能と呼ぶ。

【００５０】

図７はデータ二重化機能を利用したコントローラ相互間でＬＵの引き継ぎを行

う場合の処理動作を説明するフローチャートであり、次に、これについて説明する。ここで説明する例は、バックアップサーバとしての計算機ｎが、二重化されているＬＵの内容をテープ装置２０内にバックアップする場合のＬＵの引き継ぎの例である。

【００５１】

（１）まず、計算機２ｘは、バックアップ対象ＬＵに対するペア分割指示をディスクアレイ装置１に送る（ステップ９０１）。

【００５２】

（２）ペアが分割されると、分割されたＬＵを担当している移行元コントローラ１１ｘの構成管理手段１１２ｘは、構成情報テーブル１１２１ｘのデフォルトコントローラ番号で指定されたコントローラ１１ｎへ、対象ＬＵの構成情報を移行する。移行の際は、デフォルトコントローラ番号は、移行元コントローラの番号に変更される。また、構成情報の他に、データ二重化機能が使用するＬＵのステータスや、ＬＵの書き込み位置を記録する差分情報等の必要な管理情報も併せて移行する。移行方法は、前述した第１の実施形態の場合と同様に任意の方法により行うことができる（ステップ９０２）。

【００５３】

（３）次に、移行元コントローラ１１ｘは、データキャッシュ１１１ｘ上の対象ＬＵのデータを全てデステージする。この間、計算機２ｘからのライトアクセスは全てライトスルーとし、対象ＬＵのデータをデータキャッシュ１１１ｘ上から全て掃き出す。その際、計算機２ｘから対象ＬＵのデステージを指示するコマンドを発行してもよい（ステップ９０３）。

【００５４】

（４）デステージが完了すると、移行元コントローラ１１ｘは、移行先コントローラへデステージの完了を通知し、対象ＬＵの制御を移行先コントローラ１１ｎに切り替える（ステップ９０４）。

【００５５】

（５）対象ＬＵの制御が切り替わり、移行先コントローラ１１ｎから対象ＬＵへのアクセスが可能になると、計算機２ｎは、対象ＬＵのデータをテープ装置２０

等に格納してバックアップを取得する（ステップ９０５）。

【００５６】

（６）バックアップの取得が完了すると、計算機２ｎは、対象ＬＵに対するペア再同期指示をディスクアレイ装置１に送る（ステップ９０６）。

【００５７】

（７）ペア再同期指示を受けると、ステップ９０２の処理と同様に、対象ＬＵの制御を担当するコントローラ１１ｎの構成管理手段１１２ｎは、構成情報テーブル１１２１ｎのデフォルトコントローラ番号で指定されたコントローラ１１ｘへ、対象ＬＵの構成情報を移行する。移行の際、デフォルトコントローラ番号は、コントローラ１１ｎの番号に変更される。また、構成情報の他に、データ二重化機能が使用するＬＵのステータスや、ＬＵの書き込み位置を記録する差分情報等の必要な管理情報も併せて移行する（ステップ９０７）。

【００５８】

（８）次に、ステップ９０３の処理と同様に、対象ＬＵの制御を担当するコントローラ１１ｎは、データキャッシュ１１１ｎ上の対象ＬＵのデータを全てデステージする。この間、計算機２ｎからのライトアクセスは全てライトスルーとし、対象ＬＵのデータをデータキャッシュ１１１ｎ上から全て掃き出す（ステップ９０８）。

【００５９】

（９）デステージが完了すると、対象ＬＵの制御を担当するコントローラ１１ｎは、デフォルトコントローラ番号で指定されたコントローラ１１ｘへデステージの完了を通知し、対象ＬＵの制御を元のコントローラ１１ｘへ戻して処理を終了する（ステップ９０９）。

【００６０】

前述した本発明の第２の実施形態によれば、バックアップ処理を専用のコントローラで行うようにすることにより、通常運用系の性能に影響を与えることなくデータのバックアップを行うことが可能である。

【００６１】

図８は図７により説明したＬＵ引き継ぎの具体的な例を説明する図、図９は引

き継ぎ前後の構成テーブルの例を示す図であり、次に、図８、図９を参照して、ＬＵ引き継ぎの具体例と構成情報テーブルの書き換えとについて説明する。

【００６２】

いま、図８（ａ）に示すように、コントローラ０$^{\#}$ １１ａにＬＵ０$^{\#}$ １２１ｇ、ＬＵ１$^{\#}$ １２１ｈが割り当てられており、コントローラ１$^{\#}$ １１ｂにＬＵ２$^{\#}$ １２１ｉ、ＬＵ３$^{\#}$ １２１ｊが割り当てられているものとする。そして、図８（ｂ）に示すように、ＬＵ１$^{\#}$ １２１ｈ及びＬＵ３$^{\#}$ １２１ｊの制御をコントローラｎ$^{\#}$ １１ｎに切り替えるものとする。なお、ここで説明している例では、外部ＬＵＮ及びポート番号は変更せずそのまま使用できるものとし、また、第１の実施形態の場合と同様に、全構成管理手段１１２ｘが全ＬＵの情報を構成情報テーブル１１２１ｘに格納しているものとする。

【００６３】

前述の場合、引き継ぎ前の構成情報テーブル１１２１ｘは、図９（ａ）に示すような状態となっている。この図から、カレントコントローラ番号には、ＬＵ０及びＬＵ１に対して“０”が、ＬＵ２及びＬＵ３に対して“１”が格納されており、現在の制御を担当しているコントローラ番号を表していることがわかる。そして、ＬＵ１及びＬＵ３のデフォルトコントローラ番号は移行先コントローラであるｎが格納されている。

【００６４】

図９（ａ）に示す構成情報テーブルから移行元であるコントローラ０$^{\#}$ １１ａの構成管理手段１１２ａ及びコントローラ１$^{\#}$ １１ｂの構成管理手段１１２ｂは、デフォルトコントローラ番号で指定されているコントローラｎ$^{\#}$ １１ｎが移行先であることを知ることができる。移行の際、構成情報テーブル１１２１ｘのデフォルトコントローラ番号は移行元コントローラの番号に変更される。この結果、引継ぎ後の構成情報テーブルは、図９（ｂ）に示すようになる。このテーブルからＬＵ１及びＬＵ３のデフォルトコントローラ番号がそれぞれ“０”及び“１”に変更されていることが判る。また、引き継ぎ後の制御を担当するコントローラはコントローラｎであり、カレントコントローラ番号がｎに変更されている。

【００６５】

コントローラを切り替えてバックアップ処理が終了すると、１回目の移行とは逆向きのコントローラの移行が行われる。この場合、１回目と同様に、コントローラ$n^{\#}$ １１ｎの構成管理手段１１２ｎは、構成情報テーブル１１２１ｎのデフォルトコントローラ番号から移行先コントローラを知ることができる。これにより、自動的にＬＵの制御を元のコントローラに戻すことができる。ＬＵの制御を戻したときのコントローラとＬＵの割り当てとは図８（ａ）に示す場合と同等である。対象ＬＵの制御を戻す際、構成情報テーブル１１２１ｘの対象ＬＵのデフォルトコントローラ番号はｎに変更される。このときの構成情報テーブル１１２１ｘは、図９（ａ）に示す場合と同等であり、ＬＵ２及びＬＵ３のデフォルトコントローラ番号が再びｎに変更されていることが判る。前述したようにして繰り返しＬＵの引継ぎを行うことができる。

【００６６】

前述した本発明の第２の実施形態によれば、バックアップ処理を専用のコントローラで行うようにすることにより、通常運用系の性能に影響を与えることなくバックアップを行うことが可能となる。また、構成情報テーブルに移行すべきコントローラ番号を記録しておくことにより、自動的にＬＵの引き継ぎを行うことが可能となり、また、自動的に元のコントローラに制御を戻して、通常の処理を継続して行うことができる。

【００６７】

【発明の効果】

以上説明したように本発明によれば、各コントローラが独立した専用キャッシュを備えた場合においてもシステムを中断させることなく任意のコントローラ間で任意のボリュームを引継ぐことができ、システムにおける各コントローラの負荷分散を図ることができるという効果を奏することができる。

【００６８】

また、本発明によれば、自動的なボリュームの引継ぎ及び元のコントローラに制御を戻すことができ、これにより、バックアップ処理を専用のコントローラで行うことができ、通常運用系の性能に影響を与えることなくバックアップを行うことができる。

【図面の簡単な説明】

【図１】

本発明の第１の実施形態によるディスクアレイシステムの構成を示すブロック図である。

【図２】

構成管理手段が備える構成情報テーブルの構成例を示す図である。

【図３】

コントローラ相互間でＬＵの引き継ぎを行う場合の処理動作を説明するフローチャートである。

【図４】

図３により説明したＬＵ引き継ぎの具体的な例を説明する図である。

【図５】

引き継ぎ前後の構成テーブルの例を示す図である。

【図６】

本発明の第２の実施形態によるディスクアレイシステムの構成を示すブロック図である。

【図７】

データ二重化機能を利用したコントローラ相互間でＬＵの引き継ぎを行う場合の処理動作を説明するフローチャートである。

【図８】

図７により説明したＬＵ引き継ぎの具体的な例を説明する図である。

【図９】

引き継ぎ前後の構成テーブルの例を示す図である。

【図１０】

共有キャッシュを備えた従来技術によるディスクアレイシステムの構成を示すブロック図である。

【符号の説明】

１　ディスクアレイ装置

２ｘ（ｘ＝ａ、ｂ、……、ｎ）　計算機

３　経路制御装置

４　管理コンソール

５ｘ（ｘ＝ａ、ｂ、……、ｎ）　チャネルパス

６ｘ（ｘ＝ａ、ｂ、……、ｎ）　ＬＡＮ（Local Area Network）

７　通信手段

１０　デバイスネットワーク

１１ｘ（ｘ＝ａ、ｂ、……、ｎ）　コントローラ

１２ｘ（ｘ＝ａ、ｂ、……、ｎ）　ディスク装置

２０　テープ装置

３１　パス変更手段

４１　管理ユーティリティ

１１１ｘ（ｘ＝ａ、ｂ、……、ｎ）　データキャッシュ

１１２ｘ（ｘ＝ａ、ｂ、……、ｎ）　構成管理手段

１２１ａ～１２１ｎ　ＬＵ（Logical Unit）

１００　ディスクアレイ装置

２００ｘ（ｘ＝ａ……ｎ）　コントローラ

３００　共有キャッシュ

４００　共有メモリ

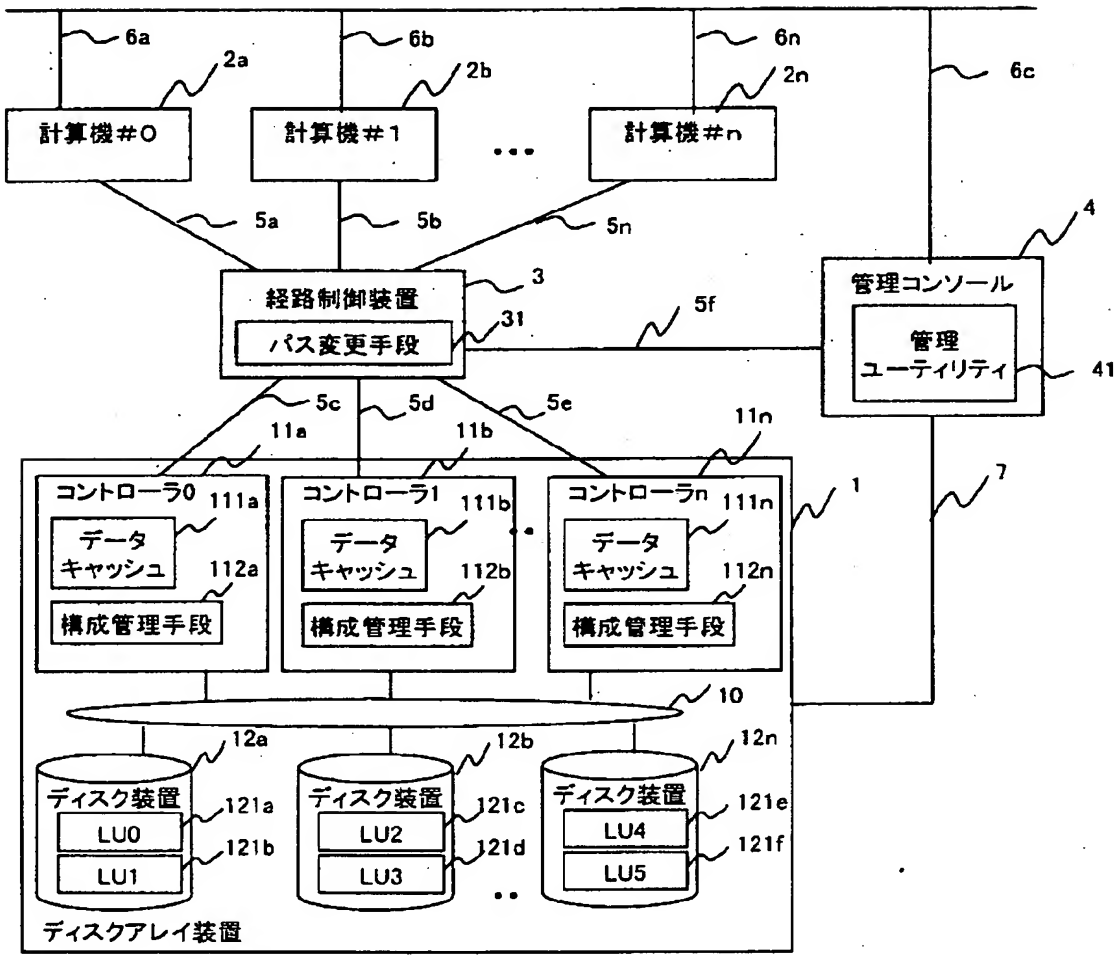
５００ｘ（ｘ＝ａ……ｎ）　ディスク装置

６００　共通バス

【書類名】　　図面

【図１】

図1


6a
6b
6n
6c
2a
2b
2n
計算機#0
計算機#1
計算機#n
5a
5b
5n
4
経路制御装置
3
パス変更手段
31
5f
管理コンソール
管理
ユーティリティ
41
5c
5d
5e
11a
11b
11n
1
7
コントローラ0
111a
データ
キャッシュ
112a
構成管理手段
コントローラ1
111b
データ
キャッシュ
112b
構成管理手段
コントローラn
111n
データ
キャッシュ
112n
構成管理手段
10
12a
12b
12n
ディスク装置
121a
LU0
121b
LU1
ディスク装置
121c
LU2
121d
LU3
ディスク装置
121e
LU4
121f
LU5
ディスクアレイ装置

【図２】

図2

構成情報テーブル

1121

内部LUN	外部LUN	ポート番号	コントローラ番号		ブロック数	RAIDGroup番号	RAIDレベル	物理アドレス情報
			デフォルト	カレント				
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

【図３】

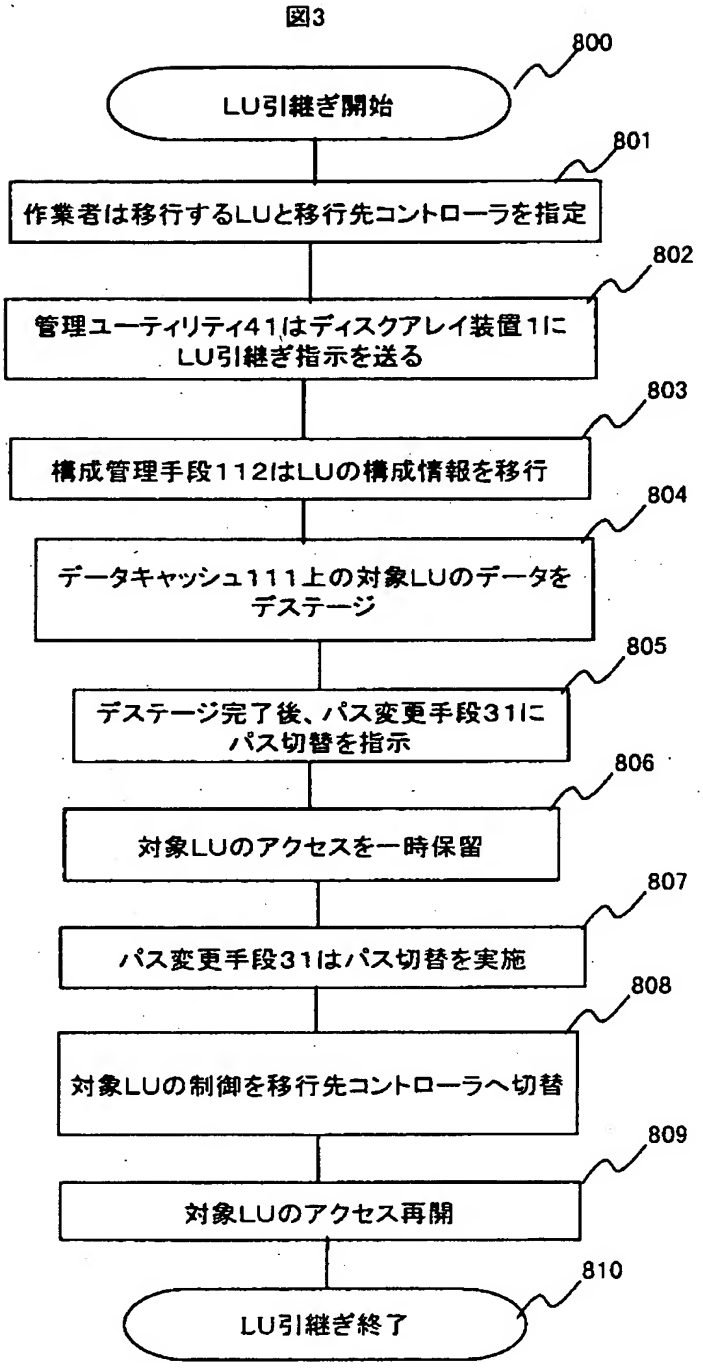
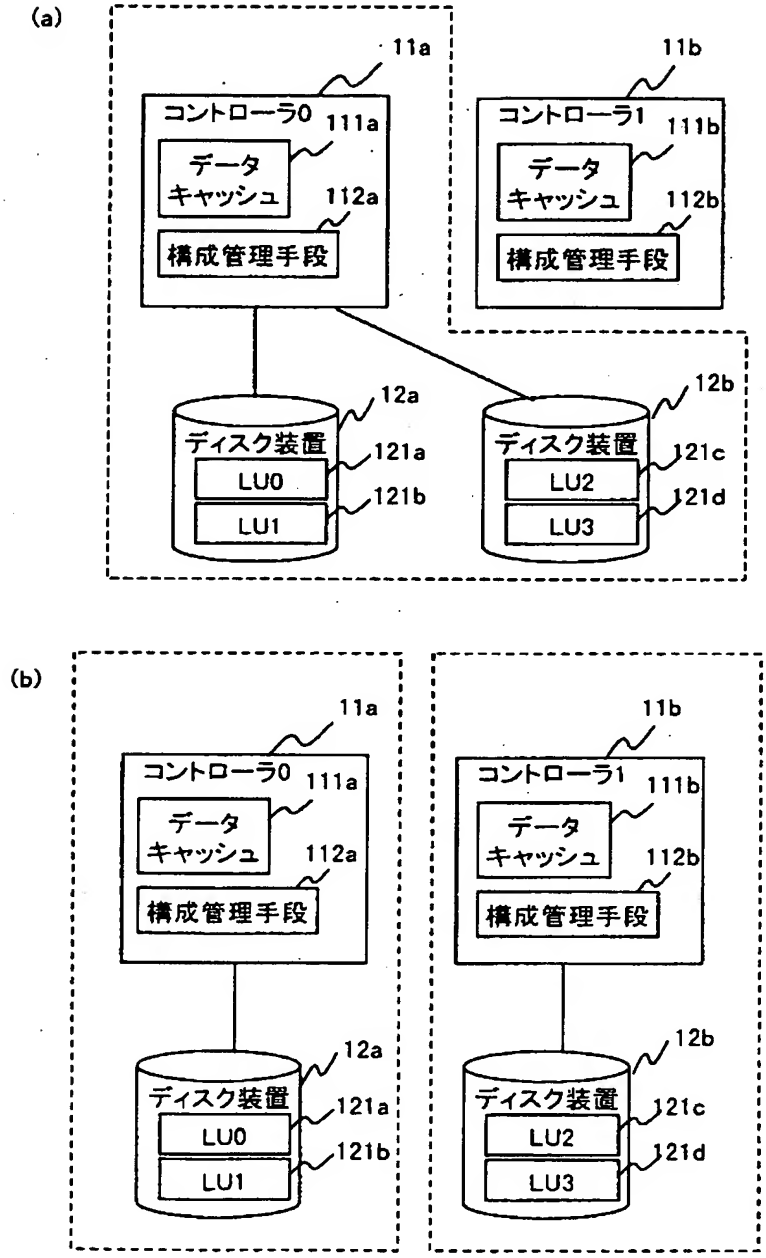


図3
800
LU引継ぎ開始
801
作業者は移行するLUと移行先コントローラを指定
802
管理ユーティリティ41はディスクアレイ装置1にLU引継ぎ指示を送る
803
構成管理手段112はLUの構成情報を移行
804
データキャッシュ111上の対象LUのデータをデステージ
805
デステージ完了後、パス変更手段31にパス切替を指示
806
対象LUのアクセスを一時保留
807
パス変更手段31はパス切替を実施
808
対象LUの制御を移行先コントローラへ切替
809
対象LUのアクセス再開
810
LU引継ぎ終了

【図４】

図4


(a)
11a
コントローラ0
111a
データ
キャッシュ
112a
構成管理手段
11b
コントローラ1
111b
データ
キャッシュ
112b
構成管理手段
12a
ディスク装置
121a
LU0
121b
LU1
12b
ディスク装置
121c
LU2
121d
LU3
(b)
11a
コントローラ0
111a
データ
キャッシュ
112a
構成管理手段
11b
コントローラ1
111b
データ
キャッシュ
112b
構成管理手段
12a
ディスク装置
121a
LU0
121b
LU1
12b
ディスク装置
121c
LU2
121d
LU3

【図５】

図5

(a)

1121x

引継ぎ前の構成情報テーブル

内部LUN	外部LUN	ポート番号	コントローラ番号		ブロック数	RAIDGroup番号	RAIDレベル	物理アドレス情報
			デフォルト	カレント				
0	1	0	0	0	⋮	⋮	⋮	⋮
1	3	0	0	0	⋮	⋮	⋮	⋮
2	0	1	0	0	⋮	⋮	⋮	⋮
3	2	1	0	0	⋮	⋮	⋮	⋮

(b)

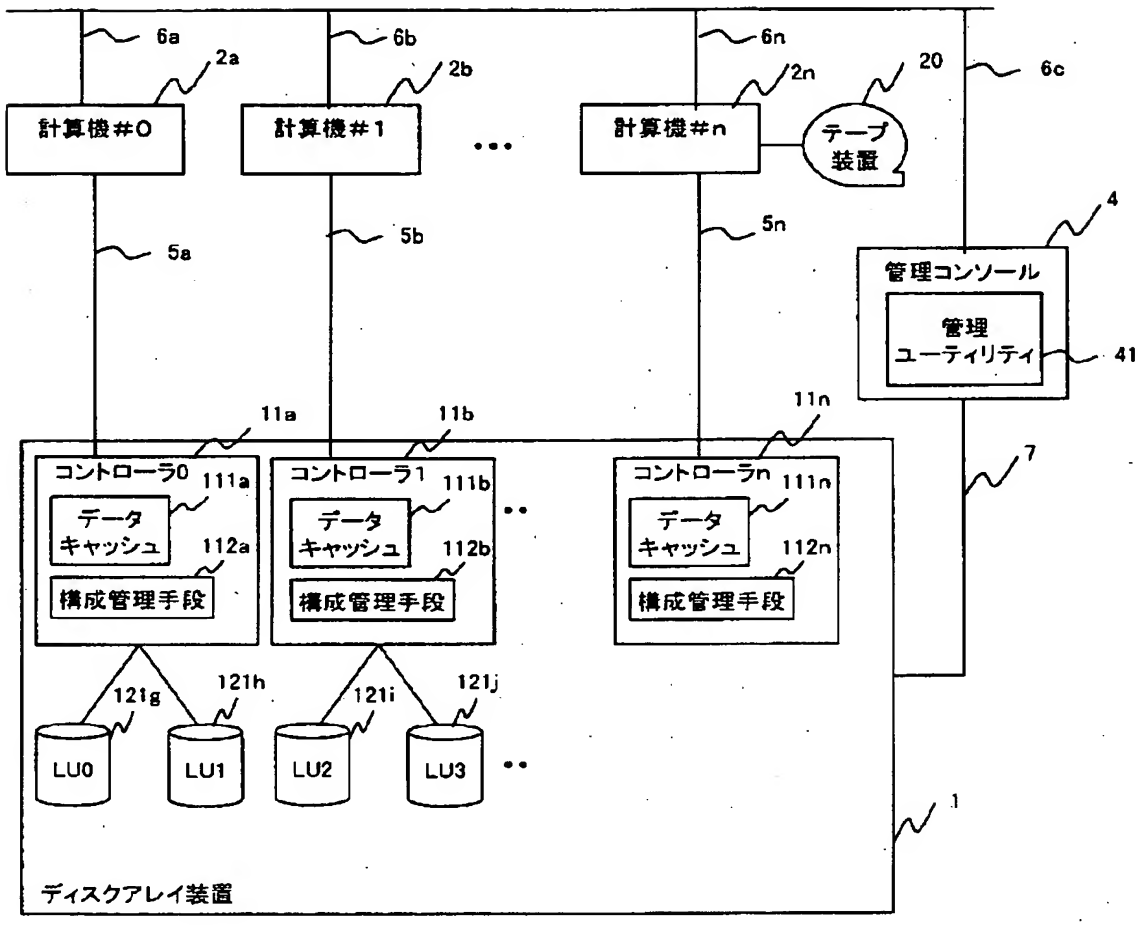
1121x

引継ぎ後の構成情報テーブル

内部LUN	外部LUN	ポート番号	コントローラ番号		ブロック数	RAIDGroup番号	RAIDレベル	物理アドレス情報
			デフォルト	カレント				
0	1	0	0	0	⋮	⋮	⋮	⋮
1	3	0	0	0	⋮	⋮	⋮	⋮
2	0	0	1	1	⋮	⋮	⋮	⋮
3	2	0	1	1	⋮	⋮	⋮	⋮

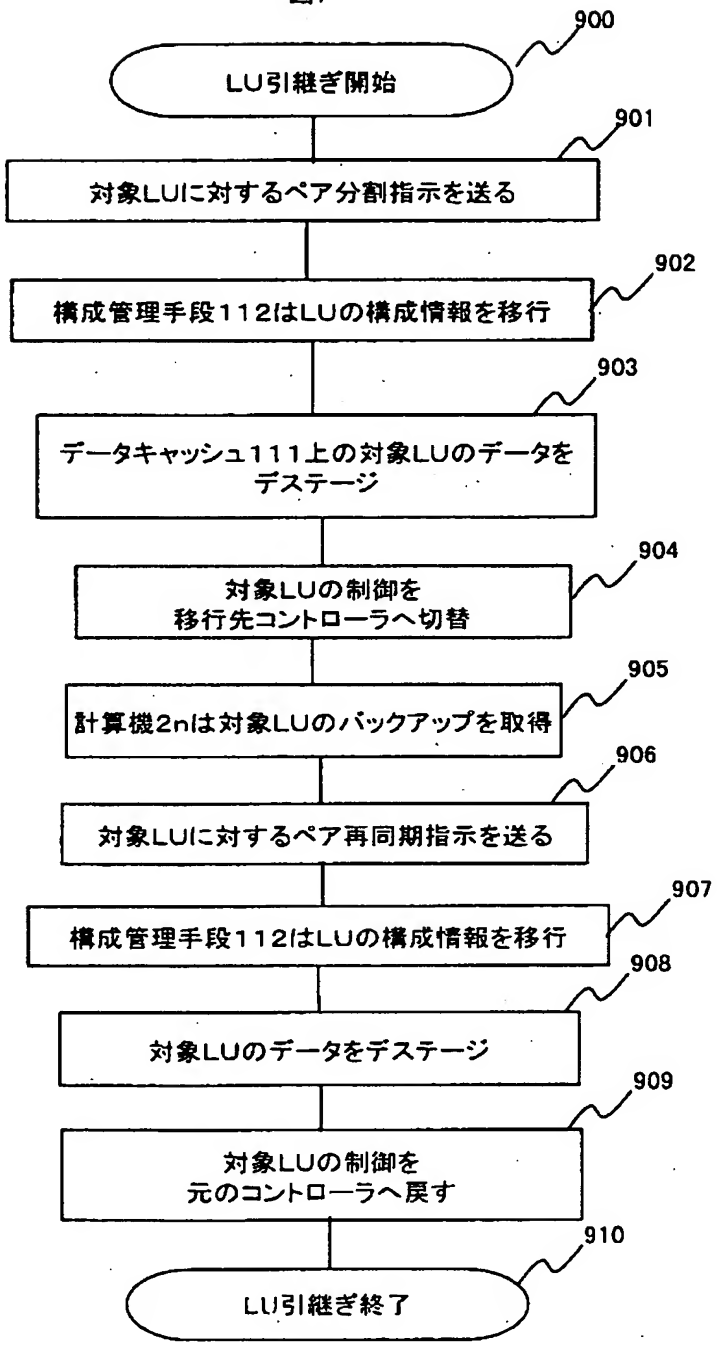
【図６】

図6


6a
6b
6n
2a
2b
2n
20
6c
計算機#0
計算機#1
計算機#n
テープ装置
5a
5b
5n
4
管理コンソール
管理ユーティリティ
41
11a
11b
11n
7
コントローラ0
111a
データキャッシュ
112a
構成管理手段
コントローラ1
111b
データキャッシュ
112b
構成管理手段
コントローラn
111n
データキャッシュ
112n
構成管理手段
121g
121h
121i
121j
LU0
LU1
LU2
LU3
1
ディスクアレイ装置

【図７】

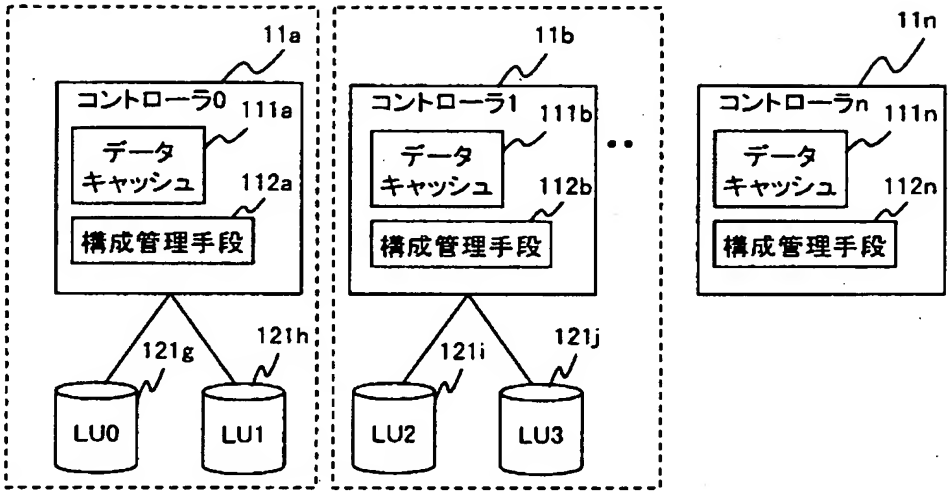
図7


900
LU引継ぎ開始
901
対象LUに対するペア分割指示を送る
902
構成管理手段112はLUの構成情報を移行
903
データキャッシュ111上の対象LUのデータを
デステージ
904
対象LUの制御を
移行先コントローラへ切替
905
計算機2nは対象LUのバックアップを取得
906
対象LUに対するペア再同期指示を送る
907
構成管理手段112はLUの構成情報を移行
908
対象LUのデータをデステージ
909
対象LUの制御を
元のコントローラへ戻す
910
LU引継ぎ終了

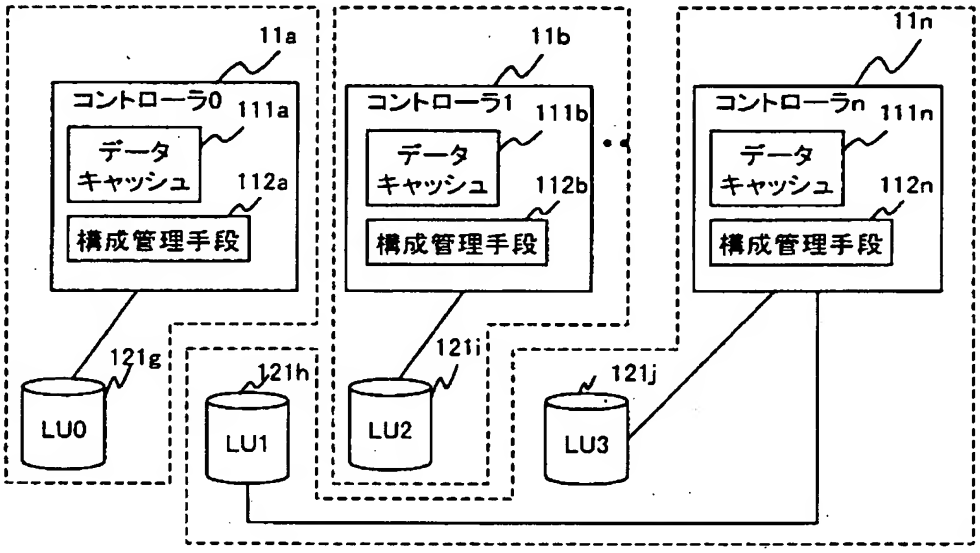
【図８】

図8

(a)


11a
コントローラ0
111a
データ
キャッシュ
112a
構成管理手段
11b
コントローラ1
111b
データ
キャッシュ
112b
構成管理手段
11n
コントローラn
111n
データ
キャッシュ
112n
構成管理手段
121g
121h
121i
121j
LU0
LU1
LU2
LU3

(b)


11a
コントローラ0
111a
データ
キャッシュ
112a
構成管理手段
11b
コントローラ1
111b
データ
キャッシュ
112b
構成管理手段
11n
コントローラn
111n
データ
キャッシュ
112n
構成管理手段
121g
121h
121i
121j
LU0
LU1
LU2
LU3

【図９】

図9

(a)

1121x

引継ぎ前の構成情報テーブル

内部LUN	外部LUN	ポート番号	コントローラ番号		ブロック数	RAIDGroup番号	RAIDレベル	物理アドレス情報
			デフォルト	カレント				
0	0	0	0	0	⋮	⋮	⋮	⋮
1	1	0	N	0	⋮	⋮	⋮	⋮
2	1	0	1	1	⋮	⋮	⋮	⋮
3	2	0	N	1	⋮	⋮	⋮	⋮

(b)

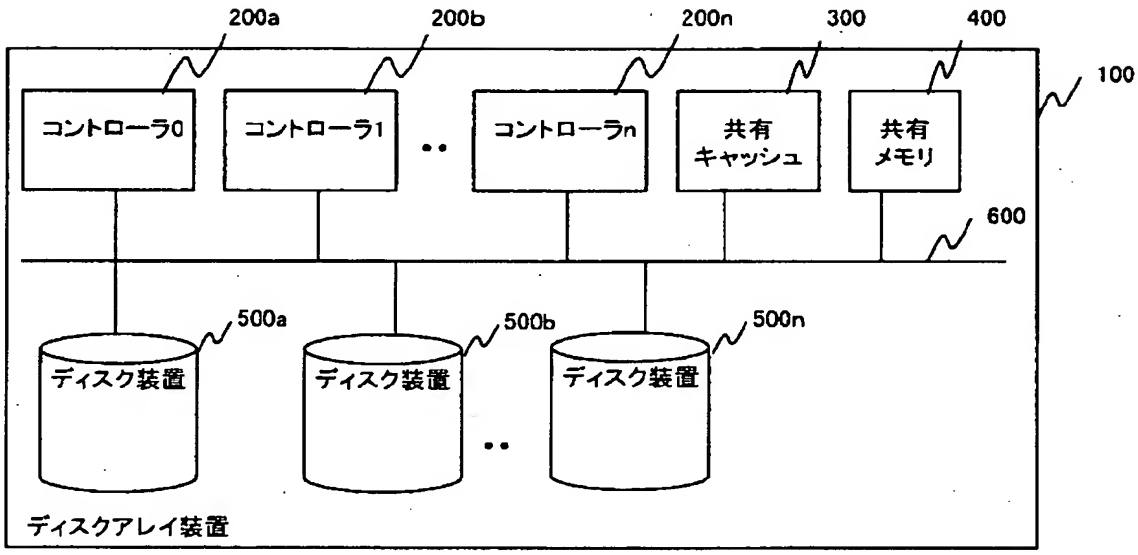
1121x

引継ぎ後の構成情報テーブル

内部LUN	外部LUN	ポート番号	コントローラ番号		ブロック数	RAIDGroup番号	RAIDレベル	物理アドレス情報
			デフォルト	カレント				
0	0	0	0	0	⋮	⋮	⋮	⋮
1	1	0	0	N	⋮	⋮	⋮	⋮
2	1	0	1	1	⋮	⋮	⋮	⋮
3	2	0	1	N	⋮	⋮	⋮	⋮

【図１０】

図10


200a
200b
200n
300
400
100
コントローラ0
コントローラ1
コントローラn
共有
キャッシュ
共有
メモリ
600
500a
500b
500n
ディスク装置
ディスク装置
ディスク装置
ディスクアレイ装置

【書類名】　　要約書

【要約】

【課題】　システムを中断させることなく任意のコントローラ間で任意のボリュームを引き継ぐことを可能とした各コントローラが独立した専用キャッシュを備えたディスクアレイシステム。

【解決手段】　コントローラ相互間でボリュームを引き継ぐとき、引き継ぎ元コントローラは、データキャッシュ上の対象ボリュームのデータをディスクに掃き出してディスクの整合性を保つようにする。これにより、各コントローラが独立した専用キャッシュを備えた場合においてもシステムを中断させることなく任意のコントローラ間で任意のボリュームを引継ぐことが可能となる。各コントローラは構成管理手段を備え、移行すべきコントローラ番号を記録しておくことにより、自動的なボリュームの引継ぎ及び元のコントローラに制御を戻すことができる。

【選択図】　　図１

出　願　人　履　歴　情　報

識別番号　　〔０００００５１０８〕

１．変更年月日　　１９９０年　８月３１日
［変更理由］　新規登録
住　所　　東京都千代田区神田駿河台４丁目６番地
氏　名　　株式会社日立製作所